BBSRC
Polaris House, North Star Avenue, Swindon, Wiltshire,
United Kingdom SN2 1UH
Telephone +44 (0) 1793 413200
Web https://bbsrc.ukri.org/

Bioinformatics and Biological Resources Fund PROPOSAL

Document Status: With Submitter
BBSRC Reference: BB/W019132/1

2021 Bioinformatics and Biological Resources Fund Committee A

Organisation where the Grant would be held

Organisation	University College London	Research Organisation Reference:	BBR Proposal
Division or Department	Sainsbury Wellcome Centre		

Project Title [up to 150 chars]

Machine Intelligence for Neuroscience Experimental Control

Start Date and Duration

a. Proposed start date	01 July 2022	b. Duration of the grant (months)	36

Applicants

Role	Name	Organisation	Division or Department	How many hours a week will the investigator work on the project?
Principal Investigator	Professor Thomas Mrsic-Flogel	University College London	Sainsbury Wellcome Centre	1.88
Co-Investigator	Professor Maneesh Sahani	University College London	Gatsby Computational Neuroscience Unit	1.88
Researcher-Co-Investigator	Dr Joaquin Rapela	University College London	Gatsby Computational Neuroscience Unit	37.5

Objectives

List the main objectives of the proposal in order of priority [up to 4000 chars]

The overall goal of this proposal is to extend, enhance, maintain and support Bonsai, a fully integrated software environment enabling cutting-edge reproducible neuroscience and behavioural science experiments, with machine-intelligence-enabled, real-time neuroinformatics methods.

Specifically, we will:

1. Develop, test and deploy new programming language interfaces to allow extending the Bonsai environment with new packages developed in languages other than C#, namely MATLAB, Python and R.

2. Define a standard API and set of operators appropriate for composing machine-learning functionality into a Bonsai workflow, specifically to enable:
2a. learning
2b. inference
2c. visualisation
2d. model-comparison and validation
2e. exchange of models and hyper-parameters

3. Incorporate a suite of machine-learning-driven analysis tools into the Bonsai environment using the standard API specified above, to implement:
3a. video-based behavioural analysis
3b. real-time interfacing of population-scale extracellular activity with laboratory control
3c. evaluation of competing data analytic approaches within a single experimental framework

4. Enhance the quality and reproducibility of systems neuroscience through:
4a. fully open-source and public-facing development
4b. rapid release of, and community engagement around, new tools
4c. provision of standard implementations of analysis and control methods, with the ability to distribute fully automated and reproducible analysis pipelines.

5. Extend and support the Bonsai user community through code, documentation, and forum engagement and development workshops.

Summary

Describe the Bioinformatic and/or Biological Resource in simple terms in a way that could be publicised to a general audience [up to 4000 chars]

Understanding the brain and the behaviour it generates is a major scientific challenge of our era. To succeed, scientists must be able to explain how animal behaviour relates to neural activity across different brain regions. This requires careful design and manipulation of behavioural experiments, where experimenters either record or manipulate neural activity while the animals (e.g. non-human primates, rodents, fish, insects) engage in specific behaviours which need to be carefully observed and quantified.

Experiments in behavioural and brain science laboratories require software that integrates and controls hardware from multiple recording devices (video, electrodes for neural activity measurement, sensors), and analysis tools that can interpret large and complex behavioural and neural datasets.

Scientists studying brain and behaviour dedicate the majority of their time designing experiments and analysing the data, with least time spent on data acquisition itself, which may impact the quality of data. Moreover, hundreds of neuroscience

research groups worldwide develop their own experimental and analytical tools, most using different programming languages, leading to inefficiencies in data sharing and analysis, and impacting reproducibility (i.e. how easy it is for someone else to repeat the same experiment).

Here we propose to provide the scientific community with a software tool that will dramatically increase the efficiency of experimental control and data analysis. We will do so by developing a new set of functionalities to an existing software platform, Bonsai. Bonsai is a fully integrated software environment that emphasises performance, flexibility, and ease-of-use, allowing scientists with no previous programming experience to quickly develop their own high-performance data acquisition and experimental control systems.

Thus far, Bonsai has been adopted by hundreds of scientists worldwide to provide interactive experimental control in behavioural and brain sciences. In this proposal, we aim to extend Bonsai's functionality with a toolbox of online and offline Machine Intelligence tools for analysis of behavioural and neural datasets, and to create an open-access platform for software sharing. Bonsai's enhanced functionality will enable new types of research, and speed up discovery and improve efficiency by (i) providing access to such tools to laboratories lacking expertise, (ii) reducing the need to reinvent the same tools in multiple labs and (iii) standardising the data processing streams, thus increasing reproducibility across laboratories. We believe this effort will enable and accelerate new discoveries in how the brain generates behaviour.

Technical Summary

Describe the Bioinformatic and/or Biological Resource in a manner suitable for a specialist reader. This summary will be made publicly available if the proposal is funded. [up to 2000 characters]

To understand the brain, scientists aim to explain how animal behaviour relates to neural activity. This requires the design and precise control of behavioural experiments, wherein animals perform particular tasks while experimenters either record or manipulate neural activity in specific neural circuits. Such experiments require data acquisition software that integrates and controls hardware from multiple recording devices (cameras, electrodes, sensors), and analysis tools that can interpret large and complex datasets. Progress is held back by the lack of standardised tools for design and implementation of experimental protocols, and the difficulty of integrating state-of-the-art data processing and neuroinformatics into custom experimental designs. The fields of behavioural and brain sciences have consequently suffered from both inefficiency and poor reproducibility, due to disparate data acquisition and analysis solutions created independently across laboratories.

To address these challenges, we propose to extend, enhance, maintain and support \textbf{Bonsai}, a fully integrated software environment to enable cutting-edge reproducible systems neuroscience experiments using animal models, with a particular emphasis on machine-intelligence-enabled, real-time neuroinformatics methods. While Bonsai is already adopted by hundreds of scientists worldwide, we aim to extend Bonsai's functionality with a toolbox of online and offline Machine Intelligence tools for analysis of behavioural and neural data (video-based analysis of behavioural motifs, real-time and offline analysis of neural signals), and create an open-access platform for software sharing and integration with multiple programming languages. Enhancing Bonsai's ecosystem will be a game-changer for behavioural and brain science experiments by enabling new types of research, increasing and diversifying user base, and dramatically improve efficiency and reproducibility of research.

Academic Beneficiaries

Describe who will benefit from this Bioinformatic and/or Biological resource(s) being funded [up to 4000 characters]

The main beneficiaries of the proposed resource will be experimental brain and animal behaviour labs worldwide who currently lack the highly technical expertise required to apply Machine Intelligence to their experiments. In fields like systems neuroscience, where progress is made by exploring the unknowns of animal behaviour and wide-scale brain physiology, the need for redesign of experimental protocols and rapid analysis of large and complex datasets is constant and critical. While increasingly-automated Machine Learning platforms are becoming available across many data science fields, they require deep expertise or training to construct and train new algorithms; or else they are easy to use but limited to very specific domains (e.g. face recognition, pose estimation). If successful, this proposal will empower an entirely new community of non-technical researchers to recombine and repurpose Machine Intelligence tools using a flexible visual programming environment and providing seamless integration with all the measurement, instrumentation and control packages already available in Bonsai.

Machine Learning and computational neuroscience experts will also benefit from a standard platform to compare and benchmark different algorithms and stress-test their tools against community datasets and experiments. The addition of new programming languages to develop and interface with Bonsai will also greatly expand the potential for application of their methods into new domains of research.

The scientific community will collectively benefit from gains in efficiency and reproducibility of deploying Machine Intelligence pipelines, and from allowing non-experts to compose existing algorithms in a flexible way to design entirely novel Machine Intelligence applications tailor-made for specific research questions. The proposed tools will generate cost savings to individual laboratories as well as funding agencies, by drastically reducing duplicated software development efforts across research groups.

In addition to the research community, the standard analysis pipelines developed in this project may be applied to other domains of society where tracking, monitoring and understanding animal behaviour would yield benefits and improved outcomes. For example, Bonsai is increasingly used in research institutions worldwide to monitor animal welfare by in-house technicians who are concerned about a broader range of environmental variables which might impact animal husbandry but which cannot be monitored in standard vivarium solutions. As a free and open-source software package, Bonsai would empower such users to prototype and design new animal monitoring systems on-site, without the need to engage in expensive software development to answer simple questions. It could thus contribute to improving the welfare of hundreds of thousands of animals in research facilities, zoos or farms. More broadly, Bonsai may also be applied to quantify dynamics of objects in wider settings from drone or satellite imagery, to track and quantify the dynamics of crop growth, deforestation or other biological ecosystems.

Summary of Resources Required for Project

Financial resources

Summary fund heading	Fund heading	Full economic Cost	BBSRC contribution	% BBSRC contribution
Directly Incurred	Staff	393728.00	314982.40	80
	Travel & Subsistence	27600.00	22080.00	80
	Other Costs	15840.00	12672.00	80
	Sub-total	**437168.00**	**349734.40**	
Directly Allocated	Investigators	40577.91	32462.32	80
	Estates Costs	50503.00	40402.40	80
	Other Directly Allocated	192.00	153.60	80
	Sub-total	**91272.91**	**73018.32**	
Indirect Costs	Indirect Costs	342563.00	274050.40	80
	Total	**871003.91**	**696803.12**	

Summary of staff effort requested

	Months
Investigator	3.50
Researcher	72
Technician	0
Other	0
Visiting Researcher	0
Student	0
Total	75.5

Animal Species

Does the proposed research involve the use of non-human primates?	Yes	✔No
Does the proposed research involve the use of dogs?	Yes	✔No
Does the proposed research involve the use of cats?	Yes	✔No
Does the proposed research involve the use of equidae?	Yes	✔No

Please select any other species of animals that are to be used in the proposed research.

Fish	Sheep
Rabbit	Rat
Amphibian	Poultry
Cow	Mouse
Reptile	Guinea Pig
Pig	Other Rodent
Bird	Other Animal

Implications

Are there ethical implications arising from the proposed research? No

Provide details of what they are and how they would be addressed [up to 4000 characters]

Other Support

Provide details of support sought or received from any other source for this or other Bioinformatic and/or Biological Resources in the same field in the past three years.

Awarding Organisation	Awarding Organisation's Reference	Title of project	Decision Made (Y/N)	Award Made (Y/N)	Start Date	End Date	Amount Sought / Awarded (£)
The Wellcome Trust Ltd	200501/Z/16/A	BonVision	Y	Y	01/05/2019	30/06/2022	50000

Staff

Directly Incurred Posts

			EFFORT ON PROJECT							
Role	**Name /Post Identifier**	**Start Date**	**Period on Project (months)**	**% of Full Time**	**Scale**	**Increment Date**	**Basic Starting Salary**	**London Allowance (£)**	**Super-annuation and NI (£)**	**Total cost on grant (£)**
Researcher-Co-Investigator	Dr Joaquin Rapela	01/07/2022	36	100	UCLGR8/37	01/08/2022	42149	3461	51902	196864
Researcher	Research Software Engineer 2	01/07/2022	36	100	UCLGR8/37	01/08/2022	42149	3461	51902	196864
									Total	393728

Applicants

Role	Name	Post will outlast project (Y/N)	Contracted working week as a % of full time work	Total number of hours to be **charged** to the grant over the duration of the grant	Average number of hours per week **charged** to the grant	Rate of Salary pool/banding	Cost estimate
Principal Investigator	Professor Thomas Mrsic-Flogel	Y	5	248	1.9	126980	19085
Co-Investigator	Professor Maneesh Sahani	Y	5	248	1.9	142994	21492
						Total	40577

Travel and Subsistence

Destination and purpose		Total £
Outside UK	RSE Conference travel and accommodation x 2	14400
Outside UK	Travel and accommodation for invited speakers	13200.00
	Total £	27600.00

Other Directly Incurred Costs

Description	Total £
Bonsai workshop x2	2400
Conference fees x 2	2640
Engagement event	2400
Laptop x 2	6000
Training courses	2400
Total £	15840

Other Directly Allocated Costs

Description	Total £
Infrastructure Technicians	192
Total £	192

Research Council Facilities

details of any proposed usage of national facilities
Research Council Facilities are not relevant to this application.

Ethical Information

Please answer the following questions as appropriate

a) Human Participation

Would the project involve the use of human subjects?		Yes	No✔
	If yes, would equal numbers of males and females be used?	Yes	No✔
Would the project involve the use of human tissue?		Yes	No✔
Would the project involve the use of biological samples?		Yes	No✔
Would the project involve the administration of drugs, chemical agents or vaccines to humans?		Yes	No✔
Will personal information be used?		Yes	No✔
	If yes, will the information be anonymised and unlinked?	Yes	No✔
	Or will it be anonymised and linked?	Yes	No✔
	Will the research participants be identifiable?	Yes	No✔

b) Animal Research

Would the project involve the use of vertebrate animals or other organisms covered by the Animals (Scientific Procedures) Act?		Yes	No✔
	If yes, what would be the maximum severity of the procedures?	Mild or non-recovery	
		Moderate	
		Severe	
	Please provide details of any areas which are Moderate or Severe:		

c) Genetic and Biological Risk

Would the project involve the production and/or use of genetically modified	Yes	

animals?			No✔
	If yes, will genetic modification be used as an experimental tool, e.g., to study the function of a gene in a genetically modified organism?	Yes	No✔
	And will the research involve the release of genetically modified organisms?	Yes	No✔
	And will the research be aimed at the ultimate development of commercial or industrial genetically modified products or processes?	Yes	No✔
Would the project involve the production and/or use of genetically modified plants?		Yes	No✔
	If yes, will genetic modification be used as an experimental tool, e.g., to study the function of a gene in a genetically modified organism?	Yes	No✔
	And will the research involve the release of genetically modified organisms?	Yes	No✔
	And will the research be aimed at the ultimate development of commercial or industrial genetically modified products or processes?	Yes	No✔
Would the project involve the production and/or use of genetically modified microbes?		Yes	No✔
	If yes, will genetic modification be used as an experimental tool, e.g., to study the function of a gene in a genetically modified organism?	Yes	No✔
	And will the research involve the release of genetically modified organisms?	Yes	No✔
	And will the research be aimed at the ultimate development of commercial or industrial genetically modified products or processes?	Yes	No✔

d) Approvals

Have the following necessary approvals been given by:			
The Regional Multicentre Research Ethics Committee (MREC) or Local Research Ethics Committee (LREC)?	Yes	No	Not required✔
The Human Fertilisation and Embryology Authority?	Yes	No	Not required✔
The Home Office (in relation to personal and project licences, and certificates of designation)?	Yes	No	Not required✔
The Gene Therapy Advisory Committee?	Yes	No	Not required✔
The UK Xenotransplantation Interim Regulatory Authority?	Yes	No	Not required✔
Administration of Radioactive Substances Advisory Committee (ARSAC)?	Yes	No	Not required✔
Other bodies as appropriate? Please specify.			

e) Other Issues

Are there any other details of which the Council should be aware?	Yes	No✔
If yes, please give details.		

Classification of Proposal

(a) Strategic Priorities

Animal health									
Bioenergy; generating new replacement fuels for a greener, sustainable future									
Collaborative research with users									
Combatting antimicrobial resistance									
Data driven biology									
Food, nutrition and health									
Healthy ageing across the lifecourse									
Integrative Microbiome Research									
International partnerships									
New strategic approaches to industrial biotechnology									
Reducing waste in the food chain									
Replacement, refinement and reduction (3Rs) in research using animals									
Research to inform public policy									
Sustainably enhancing agricultural production									
Synthetic biology									
Systems approaches to the biosciences									

Technology development for the biosciences	x								
Welfare of managed animals									

(ii) Keywords

Keyword	Research Topic	Science Area
Cognition	Systems neuroscience	Animal science
Neurocomputing	Systems neuroscience	Animal science
Neurophysiology	Systems neuroscience	Animal science
Neuroinformatics	eScience (including Bioinformatics)	Tools, technologies and methods
Decision making behaviour	Animal behaviour	Animal science
Data analysis tools	eScience (including Bioinformatics)	Tools, technologies and methods

OTHER INFORMATION

Reviewers

1	Name	Organisation	Division or Department	Email Address
Joseph Paton		Champalimaud Foundation	UNLISTED	joe.paton@neuro.fchampalimaud.org
Keywords	behavioural and neural analysis; experimental control			

Reviewers

2	Name	Organisation	Division or Department	Email Address
Dr Andreas Schaefer		The Francis Crick Institute	Research	andreas.schaefer@crick.ac.uk
Keywords	automated behavioural training, neural recordings, neural and behavioural data analysis			

Reviewers

3	Name	Organisation	Division or Department	Email Address
Dr Mackenzie Mathis		Swiss Federal Inst of Technology (EPFL)	Intelligent Systems Laboratory	mackenzie.mathis@epfl.ch
Keywords	machine learning tools; video based analysis			

Reviewers

4	Name	Organisation	Division or Department	Email Address
Dr Jakob Voigts		Massachusetts Institute of Technology	Brain and Cognitive Sciences	jvoigts@mit.edu
Keywords	open source hardware; neurophysiological data acquisition; behavioural tracking and analysis			

Proposal Classifications

Research Area:

Research Areas are the subject areas applicable to your proposal and you should select at least one of these. Once you have selected the relevant Research Area(s), please ensure that you set one as primary to facilitate the reviewer selection process. To add or remove Research Areas use the relevant link below. To set a primary area, click in the corresponding checkbox and then the Set Primary Area button that will appear.

Subject	Topic	Keyword
Animal science	Animal behaviour	

Animal science	Animal behaviour	Decision making
Animal science	Systems neuroscience [Primary]	Neurophysiology
Animal science	Systems neuroscience [Primary]	
Animal science	Systems neuroscience [Primary]	
Animal science	Systems neuroscience [Primary]	
Animal science	Systems neuroscience [Primary]	Cognition
Animal science	Systems neuroscience [Primary]	Neurocomputing
Tools, technologies and methods	Bioinformatics	Neuroinformatics
Tools, technologies and methods	Bioinformatics	

Case for support

We propose to extend, enhance, maintain and support **Bonsai**, an integrated software environment to enable **cutting-edge reproducible laboratory systems neuroscience experiments in animal models**, with a particular emphasis on **machine-intelligence-enabled, real-time neuroinformatics** methods.

1 Background: real-time neuroinformatics and control in behavioural systems neuroscience

To understand the neural basis of behaviour, experiments need to control and quantify animal behaviour while recording, and often manipulating, neural activity. Powerful experimental designs require behavioural feedback (e.g. rewards, sounds, actuated devices) or neural perturbation (e.g. using light-sensitive polarising molecules) that depend on real-time behaviour or neural state. It has also become increasingly clear that the neural signals that drive complex behaviour are manifest in a broad population of neurons rather than in single cells. Neural recording now emphasises high-dimensional signals from 100s to 1000s of neurons monitored using multi-electrode arrays or cellular-resolution imaging, with a wide array of probe configurations and neural targets. **This complexity of experiments limits the potential for database-driven progress in systems neuroscience.** The number of possible experiments is astronomical, and new questions are typically answered by designing new experiments rather than mining data from old ones.

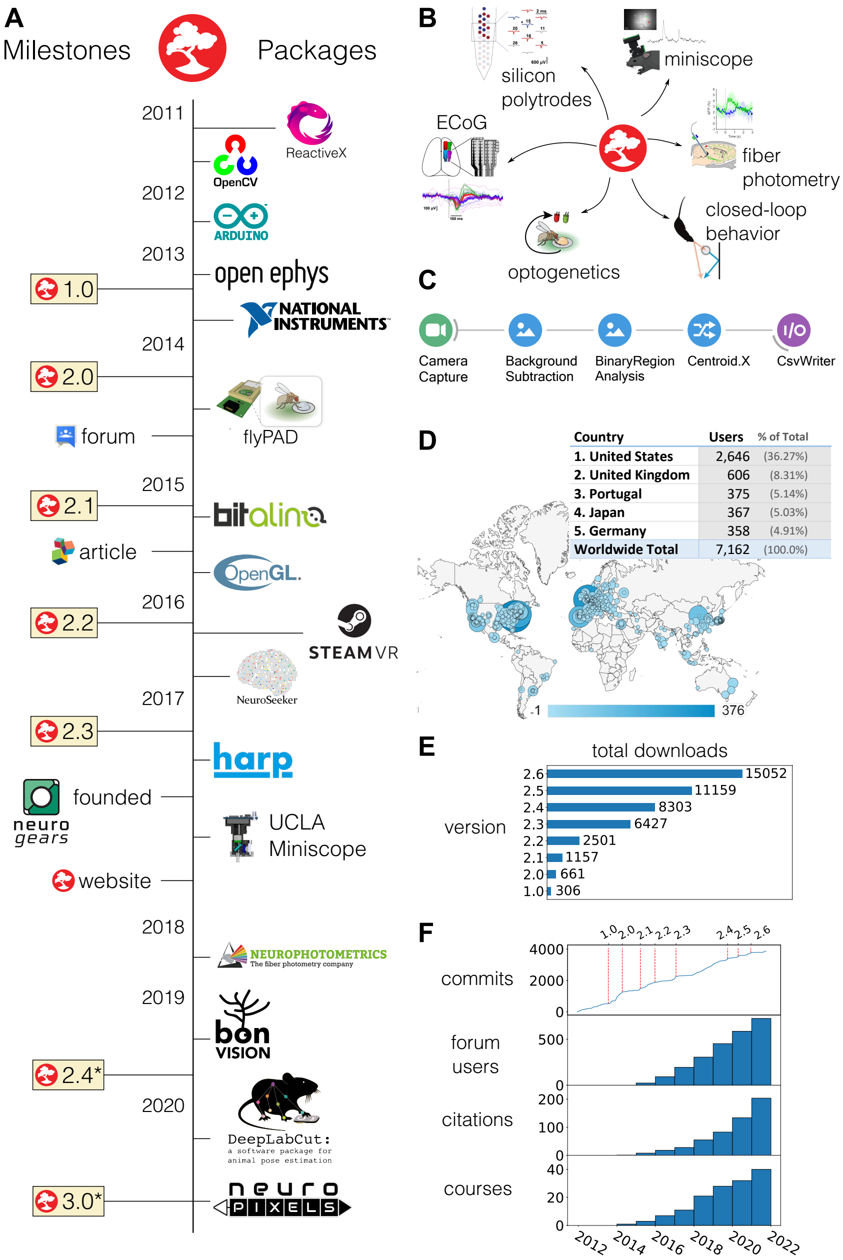


Figure 1: Bonsai development timeline and community adoption. (A) Timeline of milestones and landmark package releases. (B) Example capabilities within Bonsai. (C) Bonsai workflow for tracking the position of objects in a video. (D) Worldwide distribution of visitors to the bonsai-rx.org website. (E) Number of downloads for each Bonsai version. (F) Cumulative number of registered forum users, citations of the main publication, and number of training events, by year (bottom panels), versus release and development cycle (top panel), measured in commits to main repository.

Progress is thus held back by the challenges of designing and implementing new experimental protocols, and of integrating state-of-the-art data processing and neuroinformatics into custom experimental designs. **Reproducibility**, in particular, requires standardised tools for experimental control and analysis, shared between laboratories [1;2]. Although such tools exist in genetics, astronomy, physics and medicine [3;4;5;6;7], they are scarce in neuroscience, with documented consequences for the reliability of "discoveries" [1;8;9].

Bonsai [10;11] is a free and open-source visual programming language developed in response to these challenges. Its design emphasises performance, flexibility, and ease-of-use, allowing scientists with no previous programming experience to quickly develop their own high-performance data acquisition and experimental control systems. Bonsai combines a high-level event algebra for data streams with an integrated development environment (IDE) and an extensive library of plugins supporting multiple hardware and software packages

Case for support

used by the neuroscience research community (Figure 1A-B). A Bonsai graphical program consists of one or more source data streams (e.g. neural activity, video, audio, sensors, etc) and several interconnected operators that transform input to output datastreams (Figure 1C).

Bonsai has been adopted in hundreds of laboratories worldwide and has the largest user base in the systems neuroscience community (Figure 1D-F). In the last year alone, more than 1,000 new users incorporated Bonsai into their experimental protocols. The rapid rate of adoption of Bonsai in non-programmer experimental labs highlights the need for accessible design tools that enable state-of-the-art technology but also allow researchers to stay in control and flexibly change their experimental paradigms. Many open-source software tools are either inaccessible to non-programmers, or too constrained to be of general use outside their narrow domain of application. Bonsai has been successful because it bridges this gap.

Bonsai also supports the growing wave of foundational open hardware initiatives, such as OpenEphys[12] and UCLA Miniscope[13], allowing these tools to be quickly combined and integrated into new experiments[14]. Bonsai has been adopted in large neuroscience undertakings like the International Brain Laboratory [internationalbrainlab.com/] and the Allen Institute for Brain Science [https://alleninstitute.org/what-we-do/brain-science/]. Key to the wide adoption is the reproducibility that Bonsai ensures. Data acquisition and experimental control protocols can be replicated in any laboratory just by sharing a Bonsai configuration file.

Bonsai achieves robust behavioural control, hardware interaction and signal acquisition while ensuring remarkable ease of use. The proposed project will provide valuable enhancements to the existing codebase. We will develop, test and deploy a new software workflow making it possible to incorporate external packages into Bonsai's C#-based environment. Building on this new extensibility, we will then incorporate a suite of machine learning (ML)-driven analysis tools into the Bonsai environment. These added capabilities will allow video-based behavioural analysis as well as real-time interfacing of population-scale activity with laboratory control. In parallel we will develop the infrastructure needed to evaluate competing data analytic approaches within a single experimental framework

The proposed project will **enhance, reinforce and support** this key neuroinformatics resource to streamline **reproducible experimental design, data collection and analysis**, while making **standardised implementations of cutting-edge machine-learning methodologies** widely available to the community. This will potentially create the most powerful experimental control and analysis tool available to neuroscientists, psychologists and ethologists worldwide.

Other resources in the subject area

The large field of technologies serving experimental control and behaviour monitoring is traditionally occupied either by domain-specific graphical user interfaces for control and recording of specific devices and experiment types (e.g. Open Ephys GUI [open-ephys.org/gui/], Miniscope DAQ Software [github.com/ AharoniLab/Miniscope-DAQ-QT-Software]) or by real-time control frameworks for specifying task logic using state machine or similar formalisms (e.g. NIMH ML [monkeylogic.nimh.nih.gov/], pyControl [pycontrol. readthedocs.io/ en/latest/], Autopilot [docs.auto-pi-lot.com/en/latest/], Sanworks [sanworks.io/index.php]). These dedicated interfaces are typically very comfortable for experimenters in the specific domain for which the tool is designed, but can become unwieldy with the introduction of a new device or task variation from outside their usual scope. Alternatively, one can use a more general programming language such as Python or MATLAB, with the disadvantage of the code being harder to understand, maintain, and change. Programming languages like LabVIEW straddle the middle ground and provide a high-level, flexible visual interface for composing data acquisition and control systems. Unlike Bonsai, however, the graphical elements of LabVIEW are heterogeneous and very fine grained, thus requiring long and complex logical structures to implement even a simple experimental control system. By providing an extremely simple, yet flexible visual syntax, Bonsai provides the opportunity for even complete non-programmers to design and successfully customise relatively complex experiments from the ground up. It is this capability in particular which has made Bonsai such an attractive standard tool in experimental neuroscience.

2 Details of proposed resource enhancements

We will enhance Bonsai's extensibility in two ways. First, we will create a new package infrastructure to enable the integration of software objects implemented in non-native languages into Bonsai. Second, we will create a new general abstraction framework within Bonsai suited to the implementation and application of data analytic algorithms. Both frameworks will expose public interfaces available to all developers within the user community, thus simplifying and accelerating community-driven development.

Case for support

Furthermore, we will build on these extensibility developments to implement a suite of advanced machine intelligence neural data analysis methods that will be made available as one or more core Bonsai packages.

2.1 Extensions to incorporate non-native code packages into Bonsai

Bonsai is designed to be extensible and is distributed with a package-management framework that controls installation of optional core packages as well as user-contributed extensions. However, extensions must be written in Bonsai's native C# programming language. This restricts the range of potential contributors to the ecosystem. It also hinders incorporation of published analytic tools, many of which are implemented within other interpreted language systems such as Python, R, or MATLAB.

We will extend the Bonsai package interface to allow direct interaction with the runtime execution environments that implement these common languages. Where possible, the extensions will be based on existing pipeline tools that allow C#-based programs to interact with a number of alternative run-time engines. These include: Python.NET [github.com/pythonnet/pythonnet], R.Net [github.com/rdotnet/rdotnet] and the native .NET interface shipped with MATLAB. If necessary, we will also incorporate direct inter-runtime communication using the messaging library ZeroMQ [zeromq.org/].

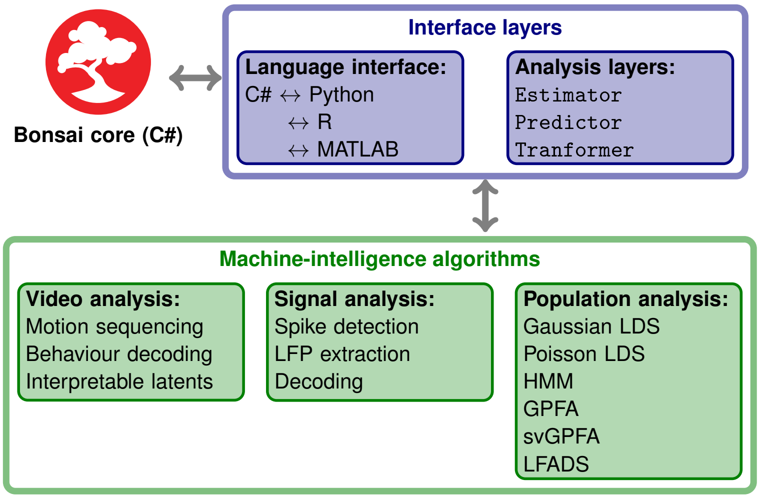


Figure 2: Proposed extensions to Bonsai

2.2 General abstraction framework for data processing

To avoid the need to re-implement a single data analysis algorithm to handle multiple different data types we will create a set of abstract machine-intelligence operators in Bonsai that can be configured to map between data substrates and ML algorithms. The planned abstraction layer is based on the successful `scikit-learn` Python library [15]. By analogy to the interfaces provided by this library we will introduce three types of Bonsai operator: `Estimator` objects for building and fitting models, `Predictor` objects for making predictions and `Transformer` objects for converting data. Interactions between these different object types will allow a single algorithm class (implemented by `Estimator` and `Predictor` operators) to act on many different data streams.

2.3 Machine intelligence functionality

We will use the frameworks developed above to integrate existing machine-intelligence programs within one or more Bonsai packages. Based on user input, we have identified initial domains of video-based behavioural analysis (Section 2.3.1) and brain-computer-interface (BCI) experiments (Section 2.3.2).

2.3.1 Video-based behavioural analysis

Precisely quantifying animal behaviour is an essential step toward understanding brain function. DeepLabCut is a Python-based software for tracking animal body parts [16], which is currently well integrated with Bonsai [17]. Here we propose extensions to Bonsai to extract other informative features from video recordings.

Motion sequencing MoSeq;[18] extracts patterns of behaviour that repeat over time (i.e. syllables of behaviour) from video data. For instance, it parses (in an unsupervised way) the behaviour of a mouse in an arena into segments of time where it was running, rearing and grooming. It uses a hidden Markov model and is implemented in Python [Code for MoSeq can be requested from the Datta laboratory, as indicated at http://datta.hms.harvard.edu/research/behavioral-analysis/.]. After trained MoSeq can be used to detect behavioural syllables online.

Decoding behaviour from neural activity BehaveNet;[19] combines hidden Markov models with convolutional autoencoders and discriminative models to decode video data from neural recordings. It is implemented in Python [https://github.com/themattinthehatt/behavenet]. After training it can be used online to decode video data and detect behavioural syllables.

Interpretable latents for behavioral videos Partitioned Subspace Variational Autoencoder, PS-VAE;[20] produces interpretable low-dimensional representations of behavioural videos by combining the output of pose-

Case for support

estimation algorithms (e.g., DeepLabCut) with unsupervised dimensionality reduction methods. These low-dimensional representations facilitate downstream behavioral and neural analyses. It is based on autoencoders and is implemented in Python [code for PS-VAE is embedded in the BehaveNet code github.com/themattinthehatt/behavenet.]. After training it can be used online to extract low-dimensional features and perform downstream processing.

2.3.2 Brain-computer-interface applications

We propose to add functionality to Bonsai for the implementation of brain-computer-interface (BCI) applications illustrated in Figure 2b. In this implementation, voltage recordings from the brain will be converted into spikes fired by single neurons by a `SpikeDetector` (Section 2.3.2.1), or to local field potentials (LFP) by an `LFPextractor` (Section 2.3.2.2). Next, neural activity (spikes or LFPs) will be represented in a low-dimensional latent space by a `LatentsCalculator`. Finally, a `Decoder` will extract the intended behaviour of the animal from this low-dimensional representation.

2.3.2.1 Spike detection

Spikes from multiple neurons can be detected in recordings from a single electrode inserted in neural tissue. Spike sorting is the computational step used to assign each spike to the neuron that generated it. Most spike sorting algorithms are designed to work offline (i.e., to use a complete recording session, after the session terminated). Online spike sorting (i.e., the task of assigning spikes to individual neurons as recordings are being acquired) is a still unsolved task, especially for recording devices with large number of electrodes. However, for the type of BCI application proposed here, previous studies have shown that spike sorting is not beneficial and high performance can be retained by just detecting spikes, without assigning them to invidivual neurons[21;22]. Thus, we will only detect spikes with a simple zero-crossing method, and not assign them to individual neurons. We will, however, integrate into Bonsai two offline spike sorting methods, Kilosort[23] written in Matlab, and MountainSort[24] written in Python, to provide Bonsai users with spike sorting functionality for applications that benefit from it.

2.3.2.2 LFP extraction

Spikes are extracted from a higher-frequency range of extracellularly recorded voltages. Local field potentials (LFPs) are another important signal to understand brain function, which is extracted from a lower-frequency range of these voltages. We propose to use in-house functions implemented in Python to compute LFPs.

2.3.2.3 Low-dimensional representations of neural recordings

Bonsai is well integrated with OpenEphys[12], which allows scientists to record neural data from a large number of devices. However, it lacks functionality to process these recordings. Here we propose Bonsai software integrations to extract interpretable summary statistics (i.e., latent variables) from neural spiking activity.

Gaussian Linear Dynamical System (GLDS)[25]: with sufficiently large bin sizes, spike counts can be modelled as Gaussian random processes. This Gaussianity assumption greatly simplifies the estimation of parameters of linear dynamical system (LDS) models, as well as inferences about their states. After models' parameters have been learned, GLDS can be used online. A unique feature of GLDS is that the posterior distribution of states given all observation up to the present can be calculated efficiently. This estimate of the posterior distribution can be used online for experimental control, as we propose in Section 2.4. We will use an R implementation of GLDS which allows the use of external inputs [https://github.com/joacorapela/kalmanFilter].

Poisson Linear Dynamical System PLDS[26]: for smaller bin sizes, spike counts are better modelled as Poisson random processes, rather than Gaussian ones. The algorithm described in [26] can estimate the parameters of a LDS, and make inferences about its states, from Poisson distributed observations. We propose to interface Bonsai with a Matlab implementation of PLDS [bitbucket.org/mackelab/ pop_spike_dyn/src/master/] that uses variational inference. PLDS does not provide online estimates of the states, since data from a full trial is needed for state inference.

Hidden Markov Model HMM;[27]: as LDSs, HMMs model a time series of observations as random processes conditioned on hidden states. However, in HMMs hidden states are discrete random variables, while in LDSs they are continuous ones. In some application domains (e.g., speech, epilepsy) discrete state assumptions are more pertinent than continuous ones. We propose to use an R implementation of HMMs [https://github.com/joacorapela/hiddenMarkovModels]. As GLDSs, once trained, HMMs can be used online to infer the posterior distribution of states given observations.

Case for support

Gaussian Processes Factor Analysis GPFA;[28]: as LDSs, GFPA models represent a time series of observation as random processes conditioned on hidden states. However, in GPFA models the state dynamics are nonlinear, while in LDS models they are linear. Thus, GPFA models are more general than LDS ones. As GLDS models, GPFA models assume that observations conditioned on states are Gaussian random processes. We propose to use a Matlab implementation of GPFA [https://users.ece.cmu.edu/ byronyu/software/gpfa0203.tgz]. GPFA models do not provide online estimates of the states, since data from the full trial are needed for state estimates.

Sparse Variational Gaussian Processes Factor Analysis svGPFA;[29]: is similar to GPFA, but models point process observations (i.e., single spikes as opposed to spike counts). We propose to use a Python implementation of svGPFA [https://github.com/joacorapela/svGPFA]. svGPFA models do not provide online estimates of states, since data from the full trial are needed for state estimates.

Latent factor analysis through dynamical systems LFADS;[30]: uses an autoencoder framework, with recursive neural networks, to infer continuous states conditioned on spike counts, similar to those inferred by GPFA. We propose to use a Python implementation at LFADS [https://github.com/tensorflow/models/tree/master/research/lfads.]. As GPFA, LFADS do not provide online estimates of states.

2.3.2.4 Low-dimensional representations of local-field potentials

Spikes are extracted from a higher-frequency range of extracellularly recorded voltages. Local field potentials (LFPs) are computed from a lower-frequency range and are another important signal to understand brain function. We propose to use states inferred from LFPs by GLDS, GPFA, HMM and LFADS (Section 2.3.2.3) as low-dimensional representations of the LFP.

2.3.2.5 Neural decoding

In order to use low-dimensional representations of spiking activity and/or of LFPs to guide the control of experiments, we need a decoder to optimally map these low-dimensional representations to experimental controls. We propose implementing in Bonsai several decoding/classification algorithms: k-nearest neighbour, linear discriminative analysis, support vector machines, random forests, artificial neural networks, naive Bayes and Gaussian process classifiers.

2.4 Comparing multiple data-analysis methods

We propose adding functionality to Bonsai to facilitate multi-method comparisons in user-supplied datasets.

2.5 Testing with neuroscience data

Multiple groups at the Sainsbury Wellcome Centre for Neural Circuits and Behaviour (SWC) currently use Bonsai for data acquisition and experimental control. Carefully testing the functionality of new software is essential to ensure correct functionality. We propose using the SWC's large Bonsai user base to test the added Bonsai functionality, before distributing it to the general public.

2.6 Reproducibility with Bonsai

Bonsai currently facilitates reproducibility of data acquisition and experimental control across laboratories. Most data acquisition and experimental control aspects of an experiment can be reproduced across different laboratories by sharing a Bonsai configuration file. Adding data analysis functionality to Bonsai will also allow this to be reproduced by Bonsai users.

We will demonstrate each machine intelligence functionality added to Bonsai (Section 2.3) and the multiple data-analysis methods comparisons (Section 2.4) in Bonsai experiments. We will then distribute Bonsai configuration files for users to reproduce these demonstrations.

2.7 Measurable targets

We propose assessing the progress of the proposed project using three types of targets. First, we will have targets associated with each release to the general public of the software components described above. Second, we will monitor user activity in the Bonsai forum [groups.google.com/g/bonsai-users], and publications citing Bonsai that use the new ML functionality, to estimate the number of Bonsai users using this functionality. Third, we will count the number of ML methods that are integrated into the Bonsai ecosystem by us and by other method developers.

3 Community demand

The need for a strong and flexible ML package that can be easily integrated with real-time experimental control workflows has been raised multiple times by the Bonsai user community, both as specific problems around tracking and interpretation of animal behaviour, and as explicit proposals to create a ML package

Case for support

[groups.google.com/g/bonsai-users/c/BZ3zOOdv_1c/m/x6OP75frBQAJ]. Even relatively modest contributions such as integrating support for specific models, such as DeepLabCut, or the more specific zebrafish feature-tracking package BonZeb, have attracted positive feedback from the community, e.g.,[17;31].

The need for an easier interface with ML algorithms is evidenced by the development of multiple graphical user interfaces for tracking animal behaviour over the last several years, especially in the much larger Python user community e.g.,[32;31]. However, these graphical user interfaces have no interaction with real-time experimental control, and their support libraries require expertise with the Python language to use flexibly. Bonsai addresses this gap, using its flexible visual programming language to combine ease-of-use and full flexibility for experimental control.

We focused this proposal on neuroscience applications of Bonsai. However, machine intelligence functionality is needed in many other experimental areas where Bonsai is or could be used. The software abstractions that will add ML functionality to Bonsai for neuroscience experiments (e.g., methods to perform multi-method comparisons) will also translate to other application domains of Bonsai.

The ongoing AI revolution makes integration of these technologies into Bonsai extremely timely. The proposed software infrastructure will facilitate the integration of novel ML methods into the Bonsai ecosystem. This will unlock new value and scientific directions from the rich behavioural and neural data that Bonsai already allows scientists to collect. It will also give non-programming experimentalists in the biological sciences the opportunity to combine real-time ML with experimental control. To our knowledge this has not been done before on such a comprehensive scale, and will, we believe, make Bonsai into a fundamental and transformative technology for the next generation of animal behavioural research.

4 User engagement

In the past four years we have developed bonsai-rx.org as a hub for documentation and learning resources (e.g. courses, video tutorials and examples) on how to use Bonsai. We have also engaged with leading, relevant training efforts such as the CAJAL Neuroscience Training Programme (Bonsai 0121) and the Transylvania Experimental Neuroscience Summer School, to increase awareness of Bonsai and its packages throughout the neuroscience community. With participation at smaller invited venues, there have been on average 5 Bonsai courses each year, with at least 20 students each; i.e. around 100 neuroscience students annually are introduced to experimental tools directly through a course in the Bonsai programming language.

Bonsai's 10th anniversary will take place in 2022. The proposed ML package will leverage planned events, and infrastructure to disseminate awareness throughout the community, including presentations at conferences, workshops, and training sessions, as well as electronic newsletters, forums and social media.

The proposed development will facilitate interfacing with other programming languages such as Python, R, and MATLAB. This will attract a whole new community of Bonsai users with expertise in software development. To provide these new users with Bonsai's full power, we plan to organise hackathons to accelerate integration of their existing data analysis and ML algorithms and methods into Bonsai using the new ML package.

Given the modular and integrative nature of the ML package, we plan to advertise its development early in the project, to allow interested partners to contribute to key components of the platform, such as the Python and MATLAB integrations. We also want to promote co-development of algorithms by early adopters, in true collaborative fashion. Given its modular nature and built-in package manager capable of handling dependency resolution and curation of community contributed content, Bonsai is well positioned for such an approach.

5 Resource management

As an open source project developed over almost 10 years, Bonsai has well-established management practices built around the user community. The source code for the core system and the majority of extension packages are hosted at `github.com/bonsai-rx`. The documentation, including learning materials and all relevant community links are publicly accessible at `bonsai-rx.org`. A community user forum is a public Google Group with more than 700 active users (Figure 1f) exchanging several new questions and answers every day. Package installation and updates are supported by standalone package management software included within the core Bonsai distribution and hosted in a shared centralised and curated repository. This ensures that bugfixes and enhancements are easily available and, importantly, allows for reproducible deployment of user environments to ensure working experiments are not disrupted by system changes. Large-scale experimental projects such as the International Brain Laboratory have leveraged this capability of Bonsai to standardise data acquisition and control software across dozens of laboratories around the world. Community technology previews are made available through the package manager, letting early adopters opt-in to the lat-

Case for support

est features before they are generally available. This allows the community to report issues with new features on GitHub, thereby distributing bug reporting and triage, and improving the quality of stable releases.

The core compiler and IDE is extensively tested, and uses a strong type system to greatly reduce bugs at compile-time. Best practices for leveraging this infrastructure for building new packages are in place.

The software engineers recruited to this project will work with our collaborators at NeuroGEARS Ltd, to strengthen and maintain these practices. Once fully familiar with the Bonsai core architecture they will participate in review of pull requests and other code contributions.

6 Long-term sustainability planning

Bonsai is a freely available open-source resource, already widely adopted by the systems neuroscience and behavioural science communities (see Fig 1). Current contributors include other large open projects (OpenEphys, DeepLabCut), and companies developing and supplying new hardware devices (Neurophotometrics). This community uptake, engagement, and commitment provides a foundation for long-term sustainability.

NeuroGEARS Ltd is a keystone of this community and a key collaborator in the current proposal. Founded by the original developer of Bonsai, it helps to support and expand the user base well beyond the life of the current proposal. NeuroGEARS currently provides quality control and curation for the core Bonsai code, and provides training and support services for research institutions around the world who want to start using Bonsai. NeuroGEARS also facilitates knowledge transfer between users and institutions adopting Bonsai, by moderating the community user forums and collaborating with organisations who want to build new innovative research projects on top of Bonsai, as in the current proposal.

The specific contributions developed as part of this project will form new core components and packages within the Bonsai ecosystem. In particular, the new language integration packages will support long-term sustainability, by dramatically increasing Bonsai's interoperability with existing software development communities which are already well established in basic science (MATLAB, Python, R). Developing these new packages with input from the core advisory panel of Bonsai, and early public distribution and user testing within the SWC and with planned workshops, will ensure solid and broad user engagement from the beginning of the development cycle and will therefore build momentum for their use, support and dissemination. Close collaboration with the core maintainers of the Bonsai ecosystem at NeuroGEARS will allow them to take over responsibility for maintenance and support of these packages beyond the end of the project, following the same principles of open-source and free sharing of knowledge that characterises the entire Bonsai ecosystem.

7 Potential for economic and societal impact

Understanding the brain and behaviour Systems neuroscientists aim to uncover how adaptive behaviour arises from computation in neural circuits across brain regions. Extending the Bonsai ecosystem will facilitate new types of experiment that causally connect neural circuits with elemental behavioural processes. Long-term, this research will help identify missing links between circuit dysfunction and behavioural deficits in neurological or mental health disorders, and will support the development of new neural circuit biomarkers and therapies. Bonsai will also enable experimental testing of emerging theories of brain function and learning, thus contributing to the neuro-inspired ML algorithms that fuel intelligent technologies, increasingly important in our society. The enhanced Bonsai platform will also support purely behavioural experiments, which track animals' posture, body parts or location in real-time, to investigate impacts of environmental variables or captivity on behaviour - crucial for animal welfare.

Improve reproducibility in behavioural and brain sciences Lack of reproducible results slows scientific progress, increasing costs to scientists and funders. Many behavioural and neural animal studies have been difficult to reproduce across laboratories. By standardising and enabling reproducible experimental design/control, as well as analysis of neural and behavioural data, Bonsai will allow multiple laboratories to implement and align experimental and analytical protocols as closely as possible.

Improve efficiency of research Most neuroscience laboratories develop custom solutions to their experimental control and analysis needs. Some laboratories also lack the technical know-how to either develop their own software for experimental hardware control or write software for state-of-the-art analysis of rich neural and behavioural data, increasingly reliant on the latest ML algorithms. The proposed upgrading of the open-source and community-empowered Bonsai ecosystem with a toolbox of online and offline analytical tools, which can dovetail with closed-loop experimental requirements, will (i) allow the deployment of such tools into laboratories that lack the necessary expertise, (ii) reduce the need to reinvent the same tools in multiple labs (cost/time efficiency) and (iii) concomitantly streamline and standardise the experimental data

Case for support

processing streams in systems neuroscience.

8 Track Record

The proposed project team comprises two academic institutes at University College London, the Gatsby Computational Neuroscience Unit (GCNU) and the Sainsbury Wellcome Centre for Neural Circuits and Behaviour (SWC), and a private collaborator, NeuroGEARS Ltd.

The GCNU is an authority in computational neuroscience and machine learning, while the SWC is a world-leading experimental neuroscience research institute. They share a common building, which facilitates regular interaction and collaboration across research groups, including the two named PIs in this project, and also collaborate in many joint activities. NeuroGEARS is a consulting company started by the creators of Bonsai. It is currently contributing to a major collaborative research project between the SWC and the GCNU. It continues to develop Bonsai, and offers courses around the globe instructing students on advanced experimental control. NeuroGEARS has used Bonsai in numerous research and private experimental projects, and has extensive experience on the implementation of neuroscience experiments.

In summary, the partners in this project have outstanding expertise on their focus areas (computational neuroscience and ML, experimental neuroscience, implementation of experiments), as well as a rich collaborative track record. The proposed project will merge this unique expertise to create a resource that will be essential for the advanced of biological sciences.

References

[1] M Baker. *Nature*, 533(26):353–66, 2016.
[2] JP Ioannidis. *Jama*, 294(2):218–228, 2005.
[3] VL Fish, K Akiyama, KL Bouman, et al. *Galaxies*, 4(4):54, 2016.
[4] H Abdalla, A Abramowski, F Aharonian, et al. *Astronomy & Astrophysics*, 612:A1, 2018.
[5] C CERN Education and O Group. 2018.
[6] ME Dickinson, AM Flenniken, X Ji, et al. *Nature*, 537(7621):508–514, 2016.
[7] C Bycroft, C Freeman, D Petkova, et al. *Nature*, 562(7726):203–209, 2018.
[8] R Botvinik-Nezer, F Holzmeister, C Camerer, et al. 2020.
[9] B KS, L JPA, M Munafò, et al. *Neuroscience*, (14):365–376, 2013.
[10] G Lopes, N Bonacchi, J Frazão, et al. *Frontiers in neuroinformatics*, 9:7, 2015.
[11] G Lopes and P Monteiro. *Frontiers in behavioral neuroscience*, 15:53, 2021.
[12] JH Siegle, AC López, YA Patel, et al. *Journal of Neural Engineering*, 14(4):045003, 2017.
[13] DJ Cai, D Aharoni, T Shuman, et al. *Nature*, 534(7605):115–118, 2016.
[14] AP Buccino, ME Lepperød, SA Dragly, et al. *Journal of neural engineering*, 15(5):055002, 2018.
[15] L Buitinck, G Louppe, M Blondel, et al. *arXiv preprint arXiv:1309.0238*, 2013.
[16] A Mathis, P Mamidanna, KM Cury, et al. *Nature neuroscience*, 21(9):1281–1289, 2018.
[17] G Kane, G Lopes, JL Saunders, et al. *bioRxiv*, 2020.
[18] AB Wiltschko, MJ Johnson, G Iurilli, et al. *Neuron*, 88(6):1121–1135, 2015.
[19] E Batty, M Whiteway, S Saxena, et al. 2019.
[20] MR Whiteway, D Biderman, Y Friedman, et al. *bioRxiv*, 2021.
[21] EM Trautmann, SD Stavisky, S Lahiri, et al. *Neuron*, 103(2):292–308, 2019.
[22] S Todorova, P Sadtler, A Batista, et al. *Journal of neural engineering*, 11(5):056005, 2014.
[23] M Pachitariu, N Steinmetz, S Kadir, et al. *BioRxiv*, p. 061481, 2016.
[24] JE Chung, JF Magland, AH Barnett, et al. *Neuron*, 95(6):1381–1394, 2017.
[25] BD Anderson and JB Moore. *Optimal filtering*. Courier Corporation, 2012.
[26] JH Macke, L Buesing, and M Sahani. *Estimating state and parameters in state space models of spike trains*, p. 137–159. Cambridge University Press, 2015.
[27] LR Rabiner. *Proceedings of the IEEE*, 77(2):257–286, 1989.
[28] BM Yu, JP Cunningham, G Santhanam, et al. *Journal of neurophysiology*, 102(1):614–635, 2009.
[29] L Duncker and M Sahani. In *Advances in Neural Information Processing Systems*, pp. 10445–10455, 2018.
[30] C Pandarinath, DJ O'Shea, J Collins, et al. *Nature methods*, 15(10):805–815, 2018.
[31] NC Guilbeault, J Guerguiev, M Martin, et al. *Scientific reports*, 11(1):1–21, 2021.
[32] T Walter and ID Couzin. *Elife*, 10:e64000, 2021.

Case for support

Justification of Resources

0.1 Directly Incurred - Staff

Research software engineers (RES)

To implement the proposed project we will recruit two experienced RSEs, appointed at grade 8. They will be responsible for developing and implementing all the work packages throughout the project. We anticipate one of the RSEs to be Dr. Joaquin Rapela, who will continue to work on integrating machine-learning behavioural and neural analysis tools with Bonsai, in collaboration with NeuroGEARS. He has the necessary machine learning and programming expertise to deliver the proposed work. An additional RSE with a broad skillset (programming in C, Python, R, Matlab, Windows) is required to deliver foreign programming language interface packages. They will work closely with the software engineering team at NeuroGEARS. Total cost to funder: £196,864 for each of the two RSEs.

0.2 Directly Incurred - Travel Subsistence

Conferences

The success of the proposed project will depend on continued training and community engagement of employed RSEs, to stay abreast of recent developments in machine learning and neural and behavioural analysis. Conferences are key resources for such knowledge acquisition. We request funding to support the attendance of at least one conference per year for each RSE. While many conferences now offer online formats, others are returning to in-person attendance, which is crucial for networking and broadening Bonsai user engagement. Possible conferences that could benefit the RSEs include SciPy, DotNet Conference, PyCon, Matlabexpo, UseR!, and SeptembRSE Conferences for Research Software Engineers.

Costs would include travel and accommodation over 3 years for each RSE. Total cost to funder: £11,520 (£5,760 per RSE).

Networking Event

We request funds for a user engagement event to mark the 10th anniversary of Bonsai. This small 3-day conference will have approximately 80-100 attendees and about 6 invited international scientists or developers who use or develop the Bonsai ecosystem. The conference will have 1 day of talks, followed by a separate 2-day hackathon (see below). The intention of this conference is to promote awareness of our software and bring together the community that uses our software to discuss possible future directions, in addition to bringing in new users. The conference would be a hybrid event with an online component for Bonsai users across the globe to also attend virtually. We request funds for travel and accommodation of invited speakers, in addition to catering for lunch and coffee breaks. We also request funds for an online platform for the conference. Travel and accommodation cost for speakers: £10,560. Catering and online platform costs: £1,920.

0.2.1 Workshops

We request funds for two Bonsai workshops to take place at the Sainsbury Wellcome Centre. The aim of the workshops is to increase dissemination and user engagement.

Bonsai for Software Developers Workshop. This will be a training course and workshop focusing explicitly on the extensibility features of the Bonsai programming languages, targeting software developers interested in making their own Bonsai packages and custom language integration. This workshop would take place at the end of year 2, towards the end of the development of the software infrastructure, as a way of presenting the development proof-of-concept to the community and possibly bringing new developers from the community to contribute to the resource from the outset.

Machine Learning in Bonsai Hackathon. This would be a collaborative workshop potentially combining scientists and engineers interested in developing applications of machine learning in Bonsai, either to address scientific questions, or to integrate new methods and algorithms into the resource (these could be developed as two alternate/complementary tracks running in parallel for participants). We expect to integrate this workshop with the user engagement event described above.

Each workshop will run over the course of two days. Registration for each workshop will be free of charge. We request funds for catering. Total cost for the two workshops: £1920.

0.3 Other Directly Incurred

Equipment

We request funds for two high-end laptop computers to facilitate possible remote working for each RSE. Total cost to funder: £4,800 (£2400 per RSE).

Training

To extend the knowledge base of RSEs working on this project, they may benefit from formal training courses in software design, programming, machine learning and/or computational neuroscience. Total cost of sign-up fees: £1,920.

Conference fees

We request funds to cover the registration fees for the two RSEs to attend at least 3 conferences for the duration of the project. Total cost to funder: £2,112 (£1,056 per RSE).

Data Management Plan

0 Proposal name

Machine intelligence for (neuroscience) experimental control

1 Description of the data

1.1 Type of study

The project will generate software modules appropriate for use in behavioural neuroscience experiments using animal models, and other types of study that involve a combination of reactive experimental (particularly behavioural) control and streamed data acquisition.

1.2 Types of data

This is a software project. We will produce and manage new modules to be integrated into the Bonsai software ecosystem. These will include extensions to incorporate non-native code packages into Bonsai (Section 2.1, Case for Support), a general abstraction framework for data processing (Section 2.2, Case for Support), machine learning functionality added to Bonsai (Section 2.3, Case for Support) and documentation.

1.3 Origin of the data

New modules will be created and tested by software research engineers at UCL in consultation with staff at NeuroGEARS Ltd (a project collaborator). These will be integrated into the existing Bonsai software ecosystem.

We will access experimental data during the testing phase, but will not seek to manage these data within the project. Such data will have been collected in the course of other projects and will continue to be managed by the relevant laboratories in accordance with the protocols they have defined.

1.4 Format and scale of the data

Software modules will comprise source files in various programming languages: C#, Python, MATLAB and R, along with supporting files for version control, compilation, testing, packaging and documentation.

2 Data management, documentation, and curation

2.1 Managing, storing and curating data

Software will be developed within the git distributed version control system, with public repositories made available within a specially created organisation at github.com and linked from `github.com/bonsai-rx`.

Mirrors of all repositories will be maintained within the Sainsbury Wellcome Centre data storage system at UCL.

Package installation and updates are supported by standalone package management software included within the core Bonsai distribution and hosted in a shared centralised and curated repository. This ensures that bugfixes and enhancements are easily available and, importantly, allows for reproducible deployment of user environments to guarantee existing experiments are not disrupted by system changes. Large-scale experimental projects such as the International Brain Laboratory have leveraged this capability of Bonsai to standardise data acquisition and control software across dozens of laboratories around the world. Community technology previews are made available through the package manager, letting early adopters opt-in to the latest features before they are generally available. This allows the community to report issues with new features on GitHub, thereby distributing bug reporting and triage, and improving the quality of stable releases.

The core compiler and the Integrated Development Environment (IDE) is extensively tested, and uses a strongly typed language that greatly reduce bugs at compile-time. Best practices for leveraging this infrastructure for building new packages are in place. The software engineers recruited to this project will work with our collaborators at NeuroGEARS Ltd, to strengthen and maintain these practices. Once fully familiar with the Bonsai core architecture they will participate in review of pull requests and other code contributions.

Data management plan

2.2 Metadata standards and data documentation

Bonsai documentation, including learning materials and all relevant community links are publicly accessible at bonsai-rx.org. Documentation will be created for the new modules produced in this project that meets the existing standards, and will be integrated into the Bonsai documentation framework.

Classes, methods and functions will be documented in the code, and automatic documentation tools (e.g., Sphinx for Python code, or Sphinx for .Net for C#) will be used to automatically generate API documentation. This documentation will be complemented with class and interaction diagrams, documenting the structure and dynamics of the application objects.

2.3 Data preservation strategy and standards

Persistent storage and long term preservation will be achieved using Zenodo [https://zenodo.org/], which is well integrated with GitHub.

3 Data sharing and access

3.1 Where will data be shared?

Software modules will be available on GitHub and on Zenodo for long-term preservation, with links from the Bonsai project site (https://bonsai-rx.org) and available through the Bonsai system's package management interface. Tutorials on how to use the added machine learning functionality will be available at the Bonsai project site (https://bonsai-rx.org) and automatically generated API documentation for this functionality will be available through ReadTheDocs [https://readthedocs.org/].

3.2 When will data be available?

Modules will be made available once they have passed rigorous internal quality checks. Software will be automatically tested during development with a comprehensive set of unit test and integration tests, implemented as GitHub Actions. Once complete, each module will be tested by a group of beta adopters (largely drawn from within the SWC to enable rapid feedback and debugging). Project outputs will be accepted when all tests run successfully and no major functional problems are reported in beta tests.

Release management: we will create a new release after each of the machine learning functionalities (described in Section 2.3 of the Case for Support) is integrated into the Bonsai ecosystem following the above method to accept a project output. Credentials for release will be granted initially by Dr. Gonçalo Lopes, Head Engineer at Neurogears and the current curator of the Bonsai official package feed, but might be expanded to other team members after training. Acceptance criteria for release will be a group decision between Dr. Lopes, Dr. Rapela, Prof. Sahani and Prof. Mrsic-Flogel.

3.3 How will data be made findable and accessible?

Bonsai is already a widely adopted software ecosystem. Community users will engage with the development of new capabilities in this project through planned workshops and the Bonsai users forum [https://groups.google.com/g/bonsai-users].

On release, modules will be announced on the Bonsai website and available for installation through the package management system.

We anticipate that incorporation of these new capabilities, associated workshops, announcements, and possible future publications will drive further adoption of Bonsai in behavioural systems neuroscience laboratories worldwide.

3.4 How will data be made reusable?

Several factors will contribute to the reusability of new functionality incorporated into Bonsai. Code implementing this functionality will be available in GitHub and Zenodo, under the permissive MIT license. Emphasis on standard and uniform code style, achieved with the help of code review and paired programming, will generate readable code, easier to read and extend by third parties. Automated build methods will simplify code building and encourage modification by new developers. Test cases with large coverage will motivate software developers to modify Bonsai, without the risk of introducing undetected bugs into the software. Carefully designed software, with simple and clear object interfaces, will simplify the reuse of the proposed software.

3.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

All distributed software will be developed as part of the project, based on published and unrestricted algorithms. Thus it is unencumbered by IP restrictions and will be available throughout development as described above.

4 Data security (where relevant)

The software produced will be freely available and open sourced. It will contain no personal data. User access to the repository will be anonymous and user personal information will not be collected. Thus security concerns do not apply.

5 Capabilities

The project collaborator, NeuroGEARS Ltd, and its lead, Dr. Gonçalo Lopes, have extended experience of software management and distribution, and will support our efforts. In addition we will benefit from advice from the UCL Research Data Management group and from the recently established UCL's Centre for Advanced Research Computing. Local storage platforms at the SWC/GCNU, totaling more than six petabytes in storage space, will be used to mirror GitHub repositories.

6 Responsibilities

Drs. Rapela (team member) and Lopes (collaborator) will jointly take responsibility for technical quality of version control, testing, documentation, and distribution. Oversight will be provided by Profs. Sahani (co-I) and Mrsic-Flogel (PI) along with a planned scientific advisory board committee that will include international experts.

7 Relevant institutional, departmental or study policies on data sharing and data security

Not applicable.

8 Author

The plan was prepared jointly by the team. Enquiries should be directed to the PI and Co-I.

Prof. Thomas Mrsic-Flogel – Narrative CV

Professor of Neuroscience, University College London (UCL)
Director, Sainsbury Wellcome Centre

Eligibility criteria: State your current position, indicating how you meet the eligibility criteria as outlined within the UKRI-BBSRC Grants Guide.

I am permanently resident in the UK and will remain resident in the UK for the duration of the proposed project.

I hold employment as a Professor with University College London. This contract of employment extends beyond the duration of the proposed grant. It is funded by a long-term grant to UCL for the core funding of the Sainsbury Wellcome Centre, provided by Gatsby Charitable Foundation and Wellcome. Funding for my professorial position is underwritten by UCL.

Contributions to the generation and flow of new ideas, hypotheses, tools or knowledge

Research in my laboratory aims to explain how the brain makes decisions by combining sensory information with previously learned knowledge. This research relies on training mice to perform complex behavioural tasks, and requires complex experimental design, combined with cutting-edge recording and data-analytic techniques, in order to produce sharable code and datasets (e.g. http://mouse.vision/han2017/) and publications (Orsolic et al 2021 Neuron; Clancy et al 2019 Nature Neuroscience; Chabrol et al 2019 Neuron; Han et al 2018 Nature; Kim et al 2018 Neuron; Khan et al 2018 Nature Neuroscience). We have also established brain-machine interfaces, wherein brain activity is used to control external actuators, requiring closed-loop experimental control (Clancy and Mrsic-Flogel, 2021 Neuron). As the behavioural tasks used in my lab require complex software-control of data acquisition and data analysis pipelines, I know first-hand their crucial importance for driving and enabling neuroscientific research.

I have published 49 peer-reviewed papers, with an h-index of 35 (calculated by google scholar).

I am a founding member of the International Brain Laboratory (IBL), a global consortium which brings together researchers from experimental and theoretical disciplines to understand decision-making in the in brain (IBL et al 2019 Neuron). The Bonsai ecosystem is critical to IBL, as it ensures that experimental control, stimulus presentation and data acquisition can be identically reproduced across all participating labs in UK, Europe and USA (IBL et al 2021, eLife)

Contributions to research teams and the development of others

As Director of the Sainsbury Wellcome Centre (SWC) at University College London, I am responsible for setting the Centre's strategic and scientific direction, which currently comprises 12 experimental labs. This includes presenting the Centre's strategy to the SWC Governing Board, including representatives of our funding organisations, to secure on-going funding. In addition, I have line management responsibilities for the Executive Team, several members of the SWC Faculty, and numerous scientific and administrative support staff.

The SWC building also houses our sister organisation, the Gatsby Computational Neuroscience Unit (GCNU; co-applicants on this proposal). Professor Sahani and I work closely together on strategic approaches to bringing experimental and theoretical neuroscience closer together. The work in this proposal would provide an additional, powerful mechanism to achieve this; it facilitates the use of the world-class Machine Learning expertise held at GCNU to contribute to driving forward experimental research at SWC and in the global neuroscience community.

As Director, I sit in the Executive Leadership Committee of the Faculty of Life Science at UCL.

As SWC's Director, I have established a Carer's support fund that makes significant financial contributions to early career researchers who require financial assistance for childcare or other dependants, thus promoting careers of individuals who would not otherwise be able to continue in active research.

In addition to this strategic responsibility, I have led a laboratory since 2008 (at UCL, as Associate Professor at the Biozentrum in Basel, and now at the SWC). I have successfully supervised 10 MA students and 7 PhD students. 8 of the 12 postdocs who have left my lab have successfully gone on to independent Group Leader positions at other institutions.

Contributions to the wider research community

In addition to the responsibilities outlined above, I am Chair of the Scientific Advisory Committee for the new Allen Institute for Neural Dynamics, which provides support and scientific direction to the Director Karel Svoboda in establishing this new long-term research area.

Since 2018, I have been a member of the ALBA steering committee, which works to promote equality and diversity in Brain Sciences in Europe. I was Chair of the FENS Programme Committee for the Berlin 2018 meeting, and a member of the FENS Executive Committee.

In response to discussions within SWC, I established the Research Culture Working Groups, which provide a platform from which to discuss and initiate actions to promote open and collaborative science, and a progressive, supportive and equitable research culture within SWC.

I acted as reviewer for numerous journals (*Nature, Nature Neuroscience, Neuron, PNAS, Journal of Neuroscience, Current Biology* and many others) and funding agencies (*Wellcome, Medical Research Council, Swiss Science Foundation, European Research Council, Royal Society, Deutsche Forschungsgesellschaft, French Science Foundation* and many others).

Contributed to broader society

I am a member of Wellcome's Mental Health Strategic Advisory Board, which helped define Wellcome's strategic approach to this major area of research and funding focus over the coming years.

I have set the strategic direction for public engagement activities at the SWC, including emphasising a focus on supporting students from underrepresented groups to successfully move into scientific study and careers. I have also contributed to public engagement activities both via SWC and via local schools in London. I have instigated a series of public lectures at the Centre to bring together members of the local community, scientists, broadcasters, philosophers and science policy makers.

1) Additional information

N/A

Prof. Maneesh Sahani – narrative CV

Academic history and eligibility criteria

I am currently **Professor of Theoretical Neuroscience and Machine Learning**, and director of the Gatsby Computational Neuroscience Unit at UCL. I am a UK resident and hold an open-ended academic employment contract with UCL.

I earned my PhD from the California Institute of Technology (Caltech) in 1999, working in the laboratories of Profs. Richard Andersen and John Hopfield. After post-doctoral work at UCL/Gatsby (with Prof. Peter Dayan) and UCSF (Prof. Michael Merzenich) I joined the academic staff of the Gatsby Unit as a Lecturer in 2004, becoming Reader in 2009 and Professor in 2013. I took over the Directorship of the Unit at the end of 2017.

Contributions to ideas, hypotheses, tools and knowledge

I have authored over 100 peer-reviewed scientific papers, with an h-index (computed by Google scholar) of 47.

A substantial component of my research focuses on the development of advance machine-learning based tools for neuroscience research. This included my PhD research. My thesis *Latent variable models for neural data analysis* introduced new techniques of probabilistic inference and applied these to the problem of "spike sorting" and clustering of neuronal activity profiles. It has been cited over 180 times.

Beginning around 2005, my group published a series of new neuroinformatics tools designed to characterise and understand population-scale activity using the large-scale multielectrode recording methods being developed. These papers provided the backbone for a new analytic approach that is now being employed and extended by systems neuroscience laboratories worldwide. A central component of the current proposal is to make this approach (and others) available within Bonsai, easing its adoption by a wider group of laboratories that lack in-house informatics expertise. Key foundational papers include:

- B. M. Yu, A. Afshar, G. Santhanam, S. I. Ryu, K. V. Shenoy, and M. Sahani. Extracting dynamical structure embedded in neural activity. In Y. Weiss, B. Schölkopf, and J. Platt, eds., Advances in Neural Information Processing Systems, vol. 18, pp. 1545–1552. MIT Press, Cambridge, Massachusetts, 2006.
  Cited 91 times.
- B. M. Yu, J. P. Cunningham, G. Santhanam, S. I. Ryu, K. V. Shenoy, and M. Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. Journal of Neurophysiology, 102:614–635, 2009.
  Introduced the GPFA method; cited 561 times.
- J. H. Macke, L. Büsing, J. P. Cunningham, B. M. Yu, K. V. Shenoy, and M. Sahani. Empirical models of spiking in neural populations. In J. Shawe-Taylor, R. S. Zemel, P. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, eds., Advances in Neural Information Processing Systems, vol. 24, pp. 1350–1358. Curran Associates, Inc., Red Hook, New York, 2011.
  Introduced effective methods to identify dynamical systems underlying population data. Cited 210 times.

This remains an active thread of my research, with key recent papers including:

- L. Duncker, G. Bohner, J. Boussard, and M. Sahani. Learning interpretable continuous-time models of latent stochastic dynamical systems. In K. Chaudhuri and R. Salakhutdinov, eds., Proceedings of the 36th International Conference on Machine Learning, vol. 97 of Proceedings of Machine Learning Research, pp. 1726–1734. PMLR, Long Beach, California, USA, 09–15 Jun 2019.
- V. Rutten, A. Bernacchia, M. Sahani, and G. Hennequin. Non-reversible gaussian processes for identifying latent dynamical structure in neural data. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H.-T. Lin, eds., Advances in Neural Information Processing Systems 33. Curran Associates, Inc., 2020.
- H. Soulat, S. Keshvarzi, T. Margrie, and M. Sahani. Probabilistic Tensor Decomposition of Neural Population Spiking Activity. Advances in Neural Information Processing Systems 34. Curran Associates, Inc., 2021.

Many of the algorithms described in these papers form the basis of capabilities to be incorporated into Bonsai as part of the proposed project.

I have also developed a number of other neuroinformatics tools relevant to the other capabiilties to be developed, and so my expertise here will be valuable for evaluation and guidance. Key examples include decoding methods:

- B. M. Yu, G. Santhanam, M. Sahani, and K. V. Shenoy. Neural decoding for motor and communication prostheses. In K. G. Oweiss, ed., Statistical Signal Processing for Neuroscience, pp. 219–263. Elsevier, 2010.

- G. Santhanam, B. M. Yu, V. Gilja, S. I. Ryu, A. Afshar, M. Sahani, and K. V. Shenoy. Factor-analysis methods for higher-performance neural prostheses. Journal of Neurophysiology, 102:1315–1330, 2009.
- B. M. Yu, C. Kemere, G. Santhanam, A. Afshar, S. I. Ryu, T. H. Meng, M. Sahani, and K. V. Shenoy. Mixture of trajectory models for neural decoding of goal-directed movements. Journal of Neurophysiology, 97(5):3763–3780, 2007.

Cell segmentation methods for optical imaging:

- M. Pachitariu, A. Packer, N. Pettit, H. Dalgleish, M. Hausser, M. Sahani. Extracting regions of interest from biological images with convolutional sparse block coding Advances in Neural Information Processing Systems 26, 1745-1753, 2013.
- G. Bohner and M. Sahani. Convolutional higher order matching pursuit. In 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), 2016.

Contributions to teams and development

To date, I have supervised a total of 14 PhDs and 14 postdocs.

Of the students, 8 now hold faculty (or group leader) positions at Cambridge, CMU, Columbia, Copenhagen, Janelia (2), Pittsburgh and Shanghai; 2 work in corporate research groups (DeepMind and Advanced Bionics); and 4 recent graduates are currently in postdoctoral positions (Aalto, Cambridge, CMU, Stanford).

Of the postdoctoral fellows, 6 have faculty positions at British Columbia, Edinburgh, Oldenburg (2), Queensland and Tuebingen; 2 hold independent research positions (Livermore Labs, Radboud); and 4 work in industry.

Since 2017, I have been Director of the Gatsby Computational Neuroscience Unit, leading strategy in research and teaching. I have created two new roles to support and develop new strategy, and have recruited two new members of faculty who have expanded our research portfolio. As Director, I sit in the Executive Leadership Committee of the Faculty of Life Sciences at UCL.

Contributions to wider research community

I have been an active contributor to the academic community, shaping the development of programme content for many scientific meetings, critically assessing the intellectual content of research proposals and awards applications, and reviewing the research and strategic direction of various world-class research institutions.

I am a member of the Society for Neuroscience and IEEE, and in 2021 became a Fellow of the European Laboratory for Learning and Intelligent Systems (ELLIS).

I served on the Board of Directors of the Computational Neuroscience Organization between 2003-2006. I was a member of the programme committees for the Neural Information Processing Systems (NeurIPS) meeting (2004,2006); and the Computational and Systems Neuroscience (COSYNE) conference (2007). I was the Workshops Chair for NeurIPS in 2008 and for the Computational Neuroscience Meeting from 1999-2003. In 2009 I was the Programme Chair for COSYNE, taking over as General Chair in 2010. I co-founded the Neural Coding, Computation and Dynamics meeting in 2007, and was in the organising committee in 2015 and 2017.

I have held editorial roles at various journals including, Current Opinion in Neurobiology (Special issue on Big Data and Neuroscience), Neural Computation, Neural Systems and Circuits, Network: Computation in Neural Systems, and Faculty of 1000, "Machine learning: life sciences" collection.

I have taken various advisory roles for international review and award panels. I reviewed the Donders Institute for Brain, Cognition and Behaviour (Netherlands), The Edmond and Lily Safra Center for Brain Sciences, Hebrew University (Israel), and the Barcelona Summer School on Advanced Modelling of Behaviour (Spain). I have advised on awards such as the Swartz Prize for Theoretical and Computational Neuroscience (SFN, US), the Bernstein Prize (BMBF, Germany), and COSYNE.

Recently, I have acted as an advisor to a new established cross-disciplinary effort at the Indian Institute of Science in Bangalore; visiting the site a number of times and hosting a joint workshop here at UCL.

I have sat on various grant award panels for the BMBF and DFG (Germany), and the (US) NSF/NIH collaborative research in computational neuroscience (CRCNS) programme. I have also acted as a reviewer for proposals for the Bernstein prize, the Einstein Foundation, EPSRC, MRC, Novo Nordisk Foundation (Denmark) and NWO (Netherlands).

Dr. Joaquin Rapela – narrative CV

Eligibility criteria

Dr. Joaquin Rapela is a Research Engineer Fellow at the Gatsby Computational Neuroscience Unit (GCNU), University College London. He has made a substantial contribution to the formulation and development of the proposed project, and will be engaged with the ensuing research. He is a Research Co-Investigator in the current application.

Contributions to ideas

Dr. Rapela has acquired unique skills for the successful implementation of the proposed project. He completed his undergraduate degree in Computer Sciences, worked as a staff software engineer at the IBM Almaden Research Center, before pursuing his doctoral degree in Electrical Engineer, with specialisation on Signal Processing, at the University of Southern California. His doctoral dissertation, as well as his two postdoctoral positions at University of California San Diego (UCSD) and at Brown University, focused on the use of advances mathematical and statistical tools to characterise electrophysiological neural signals. His publications[1] demonstrate unique skills in signal processing, statistics and computational neuroscience.

Dr. Rapela is a strong supporter of open source software, and all his publications are accompanied by free code[2]. While at UCSD, he contributed to the development of EEGLAB, a large open-source project for the characterisation of the electroencephalogram. Currently, as an Engineer Research Fellow at the GCNU, he distributes high-quality implementations of algorithms developed at unit (e.g., sparse variational Gaussian process factor analysis[3]).

Another key responsibility of Dr. Rapela at the GCNU, is to establish collaborations with experimental neuroscientists at the SWC and help them on the use of advanced statistical methods to process their state-of-the-art neural recordings.

Therefore, Dr. Rapela's expertise is perfectly suited to implement all aspects of the proposed project (neuroscience, statistics, signal processing, and software development).

Contributions to the development of others

Dr. Rapela has mentored Mr. Tsong-Yan Lin (master student at University California San Diego) on signal processing for processing electroencephalographic and electrooculographic time series [4].

Currently, Dr. Rapela is mentoring Ms. Aishah Qureshi (undergraduate student at Queens College, London), as part of the Simons Foundation Undergraduate Research Program, on the use of linear dynamical systems to understand long-range cortical communication in rodents.

Contributions to wider research community

Dr. Rapela has reviewed articles for Vision Research, Frontiers in Neuroscience and the Journal of Perceptual Imaging. He currently contributes to the SWC Research Culture Working Group [5] to improve the research culture at the SWC and GCNU.

[1] https://scholar.google.com/citations?user=eXkDg2UAAAAJ&hl=en
[2] http://www.gatsby.ucl.ac.uk/ rapela/software.htm
[3] https://github.com/joacorapela/svGPFA
[4] Joaquin Rapela, Tsong-Yan Lin, Marissa Westerfield, Tzyy-Ping Jung, and Jeanne Townsend (2012b). Assisting autistic children with wireless EOG technology. Proceedings of the 34th Annual International Conference of the IEEE EMBS, San Diego, California.
[5] https://sainsburywellcomecentre.github.io/RCWG/research

Please find attached letters of support from the following individuals:

1. **Dr Joshua Siegel**, *Assistant Investigator*, Allen Institute for Brain Science
2. **Dr Aman Saleem**, *Sir Henry Dale Fellow & Assistant Professor*, Department of Experimental Psychology, University College London
3. **Dr Tiago Monteiro**, *Research Associate*, Department of Zoology, University of Oxford
4. **Dr Marco Tripodi**, *Group Leader*, Division of Neurobiology, MRC Laboratory for Molecular Biology
5. **Dr Athena Akrami**, *Group Leader*, Sainsbury Wellcome Centre, University College London
6. **Prof. Kenneth Harris**, *Professor of Quantitative Neuroscience*, Institute of Neurology, University College London
7. **Dr Petr Znamenskiy**, *Group Leader*, The Francis Crick Institute
8. **Dr Gonçalo Lopes,** *Director*, NeuroGEARS

615 Westlake Avenue N
Seattle, Washington 98109
206.548.7000 phone

ALLEN INSTITUTE *for* BRAIN SCIENCE

Joshua H. Siegle, PhD
Assistant Investigator

joshs@alleninstitute.org
607.423.1778 mobile

November 16, 2021

Dear BBSRC BBR Grant Evaluation Committee,

I am an Assistant Investigator at the Allen Institute, where I build pipelines for high-throughput *in vivo* studies of neural activity within mouse visual system. I am also the co-founder of Open Ephys, an initiative that supports the development and dissemination of open-source tools for neuroscience research. I am writing to express my wholehearted support for the application submitted by the Gatsby Computational Neuroscience Unit, the Sainsbury Wellcome Centre, and NeuroGEARS, Ltd., with the goal of incorporating machine learning functionality into the Bonsai ecosystem.

In both of my roles, I build, interact with, and evaluate open-source software on a daily basis. Within the field of neuroscience, Bonsai stands alone in terms of its elegance, versatility, and performance. It is already used by hundreds of researchers around the world, and in my opinion it deserves to become the de facto standard tool for controlling behavior and physiology experiments. Bonsai makes it straightforward to coordinate the activity of many different asynchronous inputs and outputs in real time, while also providing interfaces to a variety of types of common hardware, including the Open Ephys acquisition board which I co-developed. The Allen Institute is in the process of migrating our visual stimulus control system to Bonsai, as a replacement for our in-house Python code, which—despite being built by professional developers over many years—was inflexible and difficult to maintain. Switching over to Bonsai has been remarkably easy, and it is substantially more performant than our previous solution.

I am extremely excited about the possibility of adding easy-to-use machine learning capabilities to Bonsai. Machine learning is widely used for offline analysis, but it is rarely incorporated into real-time experiments due to technical limitations. However, if a Bonsai user could drop in machine learning modules for classifying behavioral states and physiological events, and could use these to influence the course of an experiment in real time, it would revolutionize the way systems neuroscience experiments are carried out. An experimenter could, for example, stimulate the motor cortex only when an animal executes a particular movement—or use electrical stimulation to disrupt seizures as they are occurring. Because it already communicates with the most commonly used inputs and outputs (such as video cameras, electrophysiological amplifiers, visual displays, running wheels, lick spouts, etc.), Bonsai is the ideal platform for making machine learning tools broadly accessible for closed-loop control.

Bonsai already empowers researchers to design their experiments with an unprecedented degree of sophistication. With the addition of the proposed machine learning functionality, Bonsai will become a “must have” tool for implementing experiments that involve physiology and/or behavior. NeuroGEARS has an excellent track record of delivering robust software throughout the community, and I am very eager to start using the tools that come out of this project in my own work.

Sincerely,

Joshua H. Siegle
Assistant Investigator
Allen Institute

November 16th, 2021

Dear BBSRC panel,

I am writing to give my strongest support to the Sainsbury Wellcome Centre's BBSRC application, for funding to extend Bonsai using Machine Learning. I am an Associate Professor of Systems Neuroscience at University College London.

My laboratory uses Bonsai extensively in almost every aspect of our experiments. We use a recently developed extension of Bonsai, BonVision (Lopes et al., eLife, 2021), to present a multitude of visual stimuli from simple textures to augmented reality environments. We use Bonsai to interface and control various instrumentation such as rotary encoders and infra-red sensors. We use Bonsai to record and process videos of animals in environments, and videos of the eye. And finally, we also use Bonsai to record neural activity by interfacing with head-mounted micro-endoscopes. Bonsai is very valuable for us because it is an easy to use system with the broad capacity to deal with most aspects of our experiments, from experimental design and control to recording and online analyses.

Bonsai is already the tool of choice for many experimental neuroscience laboratories, because it allows users without significant programming knowledge to gather and process significant data streams from complex behavioural experiments. By adding Machine Learning functionality to Bonsai, the proposed work will further extend the complexity of both experimental design and data processing made possible by this system. For example, automatic identification of a specific behaviour or neural events could be the cue to adjust experimental parameters, such controlling the timing of optogenetic manipulation of neural activity in a closed-loop manner. The proposed extension of Bonsai will also increase the ease of data analysis, and generation of efficient data processing pipelines, including ensuring data can be accessed and compared by external users to promote reproducibility and transparency of data.

The extension of Bonsai by adding Machine Learning functionality will benefit my lab's research by allowing us to close the loop between our experiments recording neural activity and experiments manipulating neural activity, by giving us the ability to manipulate brain activity based on neural events. More generally, I expect that this extension will significantly expand the boundaries of what can be achieved in neuroscientific experimental design, and the depth and value of the data extracted from those experiments. For these reasons, I strongly support this proposal.

Sincerely yours,

Dr. Aman Saleem
Sir Henry Dale Fellow & Associate Professor
Department of Experimental Psychology,
University College London

University College London, Gower Street, London WC1E 6BT
Tel: +44 (0)20 7679 7511
aman.saleem@ucl.ac.uk
www.saleemlab.com

Letter of support - BBSRC Bioinformatics and Biological Resources Fund

November 16, 2021

To whom it may concern,

I am a Research Associate in the Department of Zoology at the University of Oxford. I wish to offer my full support for the application by the Sainsbury Wellcome Centre, in partnership with Neurogears, to develop the Machine Learning functionally in the Bonsai software platform.

My research focuses broady on learning, memory and decision-making of different animal organisms, such as birds, rodents, dogs and fish. For our studies we use Bonsai to run computer-based behavioural experiments, as well as to acquire, process and store different types of data such as high speed video of the animals' performance. The Bonsai ecosystem is crucial for our research as it allows for rapid and efficient prototyping and implementation of new experimental environments and paradigms, as well as real-time data processing, critical for closed-loop experimental approaches that combine task events with video tracking data. Importantly, all this is within reach of researchers without formal programming training.

I am very excited by the proposed extensions to Bonsai. Bonsai already allows users to easily and intuitively create complicated data pipelines, regardless of their previous coding experience. The proposed extension to this system will allow Bonsai to link with existing data analysis pipelines written not only in C# but in other common programming languages. This will greatly facilitate handling of the increasingly large datasets produced by my own research and behavioural research in general, and will allow us to extract even more meaningful information from any given experiment. And, by increasing the precision and depth of experimental control—for example, allowing experimental parameters to be dynamically adjusted based on either video analysis (e.g. MovSeq, BehaveNet) - the proposed extensions will allow researchers to imagine and create richer experimental designs than is currently possible.

Bonsai also already has a vibrant global user community. The proposed work will both contribute to this community by extending what is possible with this tool, and will also take great advantage from them, gathering the extensions and plug-ins that they contribute to further enrich the Bonsai ecosystem. This will help to ensure that Bonsai is sustainable, with researchers dynamically updating the open-source knowledge base.

Sincerely,

Pedro Tiago dos Santos Monteiro

Tiago Monteiro, PhD
Research Associate | Department of Zoology
University of Oxford, UK

Visiting Scientist | Domestication Lab
Konrad Lorenz Institute of Ethology
University of Veterinary Medicine Vienna, Austria

Marco Tripodi, PhD.
Neurobiology Division

Tel: (direct) +44 (0)1223 267047
Email: mtripodi@mrc-lmb.cam.ac.uk

MRC Laboratory of Molecular Biology

16th November 2021

To whom it may concern,

I wish to offer my strongest support for the development and addition of Machine Learning tools to the Bonsai software package.

Bonsai is an essential software platform in my laboratory, which focuses on understanding the neural basis of the planning and execution of goal-directed actions. We use Bonsai, a software developed by Neurogears, to control our behavioural setups and to synchronise behavioural experiments and neural recordings. This makes changing the protocols for our behavioural experiments effortless, especially for scientists without programming experience. Moreover, the Bonsai platform allows exchange of configuration files for experimental control, such that we can set up new experiments with great ease.

The proposed addition of a series of machine learning analytical tools to Bonsai will make this an incredibly powerful platform for experimental control and efficient, streamlined data analysis. We would use the video-based analysis algorithms (e.g. MoSeq, BehaveNet) to quantify and cluster behavioural motifs and their sequences. Moreover, the tools to automatically identify events in neural population data, using proposed automated spike extraction and various dynamical systems tools, will allow us to investigate which behavioural events are correlated with different neural activity measures in candidate brain regions. This would also permit a closed-loop experimental design in which identified neural or behavioural events would trigger neural activity manipulations using optogenetics, to study the causal role of particular neural populations in behaviour.

More generally, I wish to express my support for the Bonsai ecosystem. This platform provides real value to the community of behavioural and systems neuroscientists globally. It enables easy design of experiments for coding-inexperience scientists, and promotes efficiency and reproducibility of research via the transparent and shareable configuration files.

The Machine Learning extension to Bonsai would not only further enhance its functionality, but also enable scientists to deploy the same analytical tools to their data, therefore increasing the reproducibility of data analysis across laboratories. The open-source aspect of the Bonsai platform is worth highlighting, as it would create a user community to share new code for wider usage. Together, this would benefit and drive more effective progress in neuroscience in our collective endeavour to understand how neural activity gives rise to behaviour.

Best wishes,
Marco Tripodi

Marco Tripodi, PhD.
Group Leader. Division of Neurobiology
MRC Laboratory of Molecular Biology

MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH
Web: http://www2.mrc-lmb.cam.ac.uk/ - http://www.tripodilab.com

Sainsbury Wellcome Centre

25 Howland Street London W1T 4JG +44 (0)20 3108 8004 www.sainsburywellcome.org/

16th November 2021

Dear BBSRC BBR committee,
I am a Group Leader at the Sainsbury Wellcome Centre at University College London. I wish to express my strongest support for expanding the Bonsai software with machine learning tools, which will accelerate and simplify neuroscience and behavioural research worldwide.

In my lab, we employ a synergistic combination of theory and experiment to tackle the fundamental principles of neuronal computations underlying learning, inference and memory. I use the Bonsai platform for experimental control and stimulus presentation for a variety of experiments in auditory or visual domain, in freely moving or head-fixed rodents. Bonsai is an exceptionally powerful software environment and it has enabled me to rapidly and reliably design and control sophisticated experimental hardware and data acquisition. This is particularly important for scientists wishing to test new experimental protocols, expanding over several settings and devices. In particular, Bonsai, in integration with other synchronization platforms such as HARP (HArdware Research Platform) provides efficient real-time control of multiple workflows, thanks to its fast reactive language. This has enabled us to inspect animals performance, instantly, and adapt experimental paradigms in closed-loop manner

The proposal led by Professors Mrsic-Flogel and Sahani, with the industrial partner Neurogears, is exciting and transformative because it will greatly facilitate my research (and research in dozens of similar labs worldwide). Adding machine intelligence functionality to Bonsai will be increase the diversity and types of users, and make scientific experimentation in behavioural and brain sciences laboratories much easier to implement. Tools to extract behavioural parameters such as body, limb and face motion or to track animals in their environment are essential for modern neuroscience research. Coupled with algorithms for rapid extraction of neural signals (spike sorting, detection of neural events), Bonsai will enable closed-loop experimental designs that would causally probe the relationship between neural circuit activity and behaviour. These types of experiments are necessary to understand how brain function is generated by neural circuits.

Bonsai is also a platform that promotes effortless reproducibility in research via sharing of experimental protocol files and analysis packages, supported by an open-source, community-supported and Neurogears-curated platform (where users can easily contribute their own machine learning solutions to the Bonsai platform). Therefore, with the proposed enhancements, I expect Bonsai to become the main open-source environment for data acquisition and analysis, that will empower a broad community of brain and behavioural scientists to understand how the brain generates behaviour. I can only offer my strongest endorsement.

Best wishes,
Athena Akrami

ATHENA AKRAMI
Group Leader, Sainsbury Wellcome Centre (for Neural Circuits and Behaviour), Patient-Led Research Collaborative, Tel: 020 3108 8011 athena.akrami@ucl.ac.uk | https://www.sainsburywellcome.org/web/people/athena-akrami | www.lim.bio **Sainsbury Wellcome Centre** 25 Howland Street London W1T 4JG

GATSBY wellcome UCL

Letters of Support

16 November 2021

Dear BBR committee,

I am Professor of Quantitative Neuroscience at University College London and one of the founders of the International Brain Laboratory. I strongly support the extension of the Bonsai software ecosystem with machine learning functionality, as proposed in the Sainsbury Wellcome Centre's application to BBSRC's Bioinformatics and Biological Resources Fund.

The International Brain Laboratory has established a reproducible platform for behavioural experiments and neural data acquisition in mice, with Bonsai at its core. The **open-source Bonsai ecosystem** (Bonvision) is used to generate visual stimuli and enables mice to interact with the visual stimulus in a closed-loop manner. Crucially, Bonsai enables the replication of experiments across laboratories in UK, Europe and USA, via sharing a Bonsai configuration files. This reproducibility is crucial to IBL's progress in relating decision-making to activity in neural circuits.

Adding machine intelligence functionality to Bonsai will be a transformative addition to what is already a very powerful package. Machine learning is transforming science but it has yet to have a profound impact in experimental control. The proposed extension to Bonsai could make this happen, with important implications for neuroscience research. For example, the new machine intelligence functionality will allow Bonsai users to control experimental parameters or interventions based on automatic extraction of either behavioural

Kenneth D. Harris, PhD
Professor of Quantitative Neuroscience
UCL institute of Neurology and UCL Department of Neuroscience, Physiology & Pharmacology
21 University Street, London, WC1E 6DE, United Kingdom
kenneth.harris@ucl.ac.uk

motifs (e.g. different postures or movements) or neural activity events in different brain regions. These new variables for experimental control would revolutionize behavioural and systems neuroscience by allowing new types of closed-loop experiment that would have been inconceivable even a few years ago.

The Bonsai platform allows users with only basic programming expertise to design complex behavioural experiments with ease, and enables replication by allowing them to share the config files with the rest of the community. The integration of machine learning tools to control experiments and to analyse behavioural and neural data on- and off-line (video-based analysis of behaviour, online spike sorting, low-dimensional representation of neural data, neural decoding) will accelerate discovery by enabling the design of new types of experiment, and will also speed up data analysis once experimental data have been acquired. The ability for users to contribute their own machine learning code/algorithms to the Bonsai ecosystem will also increase user engagement and empower the neuroscience community via this open-source platform. I am delighted to offer this project my strongest support.

Sincerely,

Kenneth D. Harris

Professor of Quantitative Neuroscience

The Francis Crick Institute Limited
1 Midland Road, London, NW1 1AT
+44(0)203 7960000
info@crick.ac.uk
www.crick.ac.uk

Dr. Petr Znamenskiy
Group Leader
petr.znamenskiy@crick.ac.uk

16 November 2021

Dear selection panel,

I would like to offer my strongest support for the application for BBSRC Bioinformatics and Biological Resources funding to support the extension of the Bonsai programming ecosystem using Machine Intelligence. The Sainsbury Wellcome Centre's proposal will significantly increase the value of this resource for my own research and that of the wider neuroscientific community.

I am a Group Leader at the Francis Crick Institute focusing on understanding the organisation of neural circuits in the visual system and their role in processing visual signals in actively moving animals. In my research, we use Bonsai as the primary tool for experimental control, visual stimulation and recording of animal behaviour. Bonsai is a very valuable resource because it seamlessly integrates with a wide range of experimental hardware and enables us to easily create virtual reality environments, where the visual stimulus is updated in closed-loop based on animals' movements with short latencies.

Bonsai allows researchers to design and implement complex behavioural experiments. Bonsai's flexibility and the extent of its adoption is demonstrated in the wide array of plug-ins and extensions developed by the global user community. This will be further extended by the proposed Machine Learning algorithms for video-based and neural analysis, which will allow users to integrate Bonsai with new and existing data analysis methodologies written in multiple programming languages. Furthermore, the proposed enrichment of Bonsai with video behavioural analysis, integration with Brain-Computer-Interfaces, and the ability to decode behaviour from neural activity will all not only facilitate current research, but will expand the boundaries of experimental design.

The proposed work brings together the creators of Bonsai (Neurogears) with researchers with both experimental and machine learning expertise; an ideal team to ensure that the extended package is designed to precisely meet the needs of the neuroscientific and wider scientific communities. I wholeheartedly support this proposal for BBSRC funding.

Sincerely,

Petr Znamenskiy

NeuroGEARS Ltd
31 Oval Road
London NW1 7EA
United Kingdom

17th November 2021

Biotechnology and Biological Sciences Research Council
Polaris House
North Star Avenue, Swindon
Wiltshire SN2 1UH
United Kingdom

Dear BBSRC BBR grant evaluation committee,

This letter confirms the intention of NeuroGEARS Ltd to participate in the project proposed by Prof. Thomas Mrsic-Flogel and collaborators. I have created NeuroGEARS as a technology company bridging Neuroscience, Games, Interaction, and Robotics. We engage with top research and academic institutions worldwide to reimagine science by using and creating open-source tools which are accessible, understandable, and modifiable. Our core mission is to make scientific knowledge available to everyone, and to empower people to follow their curiosity, wherever it may lead.

NeuroGEARS employs the core team behind the visual reactive programming language Bonsai, which now powers thousands of experiments around the world. Bonsai is at the centre of a vibrant and growing ecosystem of hardware and software projects in Neuroscience, including Open-Ephys, DeepLabCut, Miniscope, Harp, and Neurophotometrics.

Our strong partnership with academia has allowed us to multiply the impact of our activities and mobilize and engage with scientists across a variety of fields. Being based in London, we have pursued several collaborations with local, national, and international organisations to develop new open-source technologies, and specifically we focus on the long-term sustainability and support of these platforms to enable broad accessibility of advanced features to non-technical researchers. These collaborations also give us a privileged view into the wider needs of the community, and we have noticed recently increasing requests for machine learning functionality and sophisticated behaviour analysis in the development projects we undertake, such as extracting movement activity patterns to classify the intentional state of the animal or predict its next action to cancel out latency in closed-loop stimulation.

Letters of Support

Great technological advances frequently occur when scientists from different areas join forces in a common effort, as in this collaboration between Bonsai, the Gatsby Computational Neuroscience Unit (GCNU) and the Sainsbury Wellcome Centre (SWC). The GCNU is a world-leader in the development of advanced methods for the characterization of neural signals. The addition of these methods to Bonsai will enable it to control experiments and analyse recordings in unprecedented new ways. The SWC is performing state-of-the-art neuroscience experiments and it will provide excellent test cases to evaluate the methods advanced in this proposal. I have myself concluded my PhD in Neuroscience at the SWC in the lab of Dr. Adam Kampff and am currently personally coordinating an unprecedented large-scale collaboration across labs at the Centre which will provide a unique opportunity to test the machine learning functionality advanced by this proposal in the field of complex and continuous multi-animal tracking and dense extracellular neurophysiology. Given my past, present and future experience with the needs and challenges of analysing animal behaviour, I am confident that the methods we propose to develop will benefit any scientific discipline that relies on experimental control and monitoring of animal behaviour.

A unique feature of this proposal is that it does not aim just to implement advanced machine learning methods in Bonsai. Instead, it offers to create generic software infrastructure to facilitate the integration of existing, and yet to be developed, machine learning algorithms into the Bonsai ecosystem and enhance interoperability with popular programming ecosystems like MATLAB, Python, and R. This is very much in line with the openness of Bonsai, whose current popularity in in part due to its ability to interact with many other open-source tools.

I hence express our complete enthusiasm for participating in the proposed project, which promises to add to Bonsai machine learning capabilities which will drastically augment its functionality, transform its community of users, and allow us to expand our mission of empowering scientific experiments in entirely new directions.

Dr. Gonçalo Lopes

Director, NeuroGEARS Ltd

25 Howland Street London W1T 4JG +44 (0)20 3108 8004 www.sainsburywellcome.org/

17th Nov 2021

To whom it may concern,

Please find attached our BBSRC Bioinformatics and Biological Resources application titled "*Machine intelligence for (neuroscience) experimental control*". The current proposal requests funding to maintain and enhance an existing resource with additional machine learning functionality. We declare no conflicts of interest.

Sincerely,

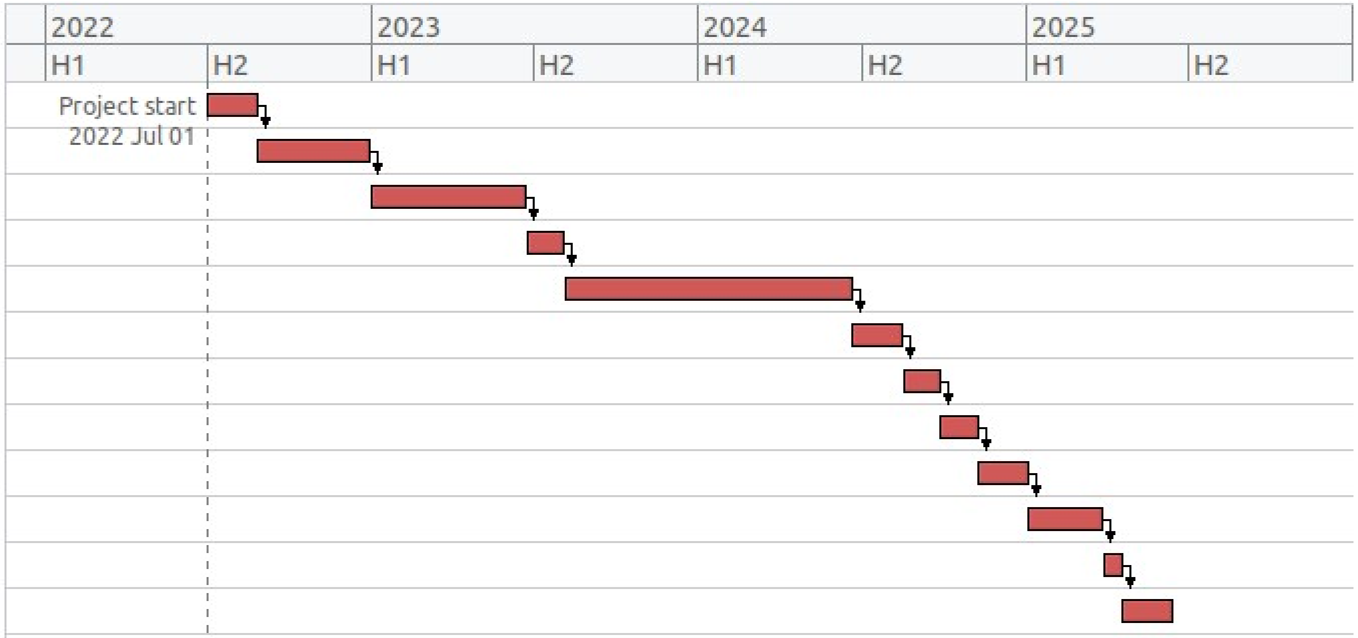
Tom Mrsic-Flogel
Professor of Neuroscience and Director of the Sainsbury Wellcome Centre

Maneesh Sahani
Professor of Theoretical Neuroscience and Machine Learning, Director of the Gatsby Computational Neuroscience Unit

Dr Joaquin Rapela
Research Associate, Gatsby Computational Neuroscience Unit

Proposal cover letter.

Name	Duration
RSE training	40d
2.1 Software infrastructure	90d
2.2.1 Video-based behavioral analysis	125d
2.2.2.1 Spike detection and sorting	30d
2.2.2.2 Low-D repres. of spiking data	230d
2.2.2.3 Low-D repres of LFPs	40d
2.2.2.4 Neural decoding	30d
2.2.2.5 Comparing multiple methods	30d
Integration testing	40d
Documentation	60d
Conference attendance	15d
Workshops preparation	40d



An overview project management and proposed scientific advisory board